Neurocomputing 219 (2017) 88–98

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Convolutional neural networks for hyperspectral image classification

Shiqi Yu [a], Sen Jia [a,*], Chunyan Xu [b]

[a] College of Computer Science and Software Engineering, Shenzhen University, China
[b] School of Computer Science and Technology, Nanjing University of Science and Technology, China

A R T I C L E I N F O

Article history:
Received 8 July 2016
Received in revised form
30 August 2016
Accepted 8 September 2016
Communicated by Jiwen Lu
Available online 13 September 2016

Keywords:
Hyperspectral image classification
Convolutional neural networks
Deep learning

A B S T R A C T

As a powerful visual model, convolutional neural networks (CNNs) have demonstrated remarkable performance in various visual recognition problems, and attracted considerable attention in recent years. However, due to the highly correlated bands and insufficient training samples of hyperspectral image data, it still remains a challenging problem to effectively apply the CNN models on hyperspectral images. In this paper, an efficient CNN architecture has been proposed to boost its discriminative capability for hyperspectral image classification, in which the original data is used as the input and the final CNN outputs are the predicted class-related results. The proposed CNN infrastructure has several distinct advantages. Firstly, different from traditional classification methods those need hand-crafted features, the CNN model used here is designed to deal with the problem of hyperspectral image analysis in an end-to-end way. Secondly, the parameters of the CNN model are optimized from a small training set, while the over-fitting problem of the neural network has been alleviated to some extent. Finally, in order to better deal with the hyperspectral image information, 1×1 convolutional layers have been adopted, and an average pooling layer and larger dropout rates have also been employed in the whole CNN procedure. The experiments on three benchmark data sets have demonstrated that the proposed CNN architecture considerably outperforms other state-of-the-art methods.

1. Introduction

With the rapid development of remote sensors, the acquisition and collection of hyperspectral data has become much easier and more affordable, making hyperspectral image analysis to be one of the most promising techniques in many practical applications, including precision agriculture, environmental monitoring, military surveillance, etc. [1,2]. Hyperspectral image (HSI) data often contains hundreds of spectral bands over the same spatial area, which has provided valuable information to identify the various materials [3,4]. HSI classification is similar with image labeling in computer vision field [5]. One difference between them is the number of data bands. There are normally about 200 bands for a HSI, but only 3 or 4 ones in image labeling. Another difference is the number of labeled sample. For image labeling, it is relatively easy to label samples. In HSI classification, due to the difficulty and expense of manually labeling, the limited availability of labeled training samples is the main obstacle of hyperspectral image classification. Hence the small sample set (3S) problem has attracted increasing attention in recent years [6,7].

One reasonable way to tackle the 3S problem is dimensionality reduction [8,9], which could largely reduce the impact of Hughes phenomenon, i.e., a large amount of labeled samples is needed for the high-dimensional data to obtain reliable results [10]. Dimensionality reduction can be accomplished by transforming the original hyperspectral data into a low-dimensional space (referred as feature extraction) [11,12], such as principal component analysis (PCA) [13,14], independent component analysis (ICA) [15], manifold learning [16] or directly picking out the most representative bands from the hyperspectral data (referred as band selection) [17,18], such as the ranking-based methods [19,20] and clustering-based methods [8,21,22]. But some important information may be lost during the dimensionality reduction process. More severely, the features obtained through dimensionality reduction in the spectral domain cannot fully characterize the properties of the materials, hence more discriminative features should be extracted.

Fortunately, spatial information, which reflects the fact that the adjacent pixels in the spatial domain belong to the same class with a high possibility, is a valuable complement to the spectral signatures, and has been extensively studied for hyperspectral image classification [23]. Specifically, the mathematical morphology method applies the opening and closing morphological transforms on several principal components to obtain features containing spatial structure information [24]. Other spatial filters [25,26] are

* Corresponding author.
E-mail addresses: shiqi.yu@szu.edu.cn (S. Yu), senjia@szu.edu.cn (S. Jia), xuchunyan01@gmail.com (C. Xu).

http://dx.doi.org/10.1016/j.neucom.2016.09.010
0925-2312/

also used to exploit the spatial regularity of materials. Moreover, the contextual information can be used to refine the classification results through a regularization process in the postprocessing stage [27]. However, a large number of training samples are generally required to adequately characterize the large variability of the objects, which is difficult to meet in practice. Alternatively, in order to extract the joint spatial-spectral features, three-dimensional wavelet-based methods, especially the Gabor filters, have been proposed to simultaneously fuse the spectral and spatial information, which has shown competitive classification performance for the 3S problem [28–30]. Recent advances have revealed that Gabor filters with different predefined orientations and scales are a kind of convolutional filters, whereas the popular convolutional neural networks (CNNs) can learn convolutional filters automatically [31]. These encouraging results have motivated us to apply the CNN model for hyperspectral image classification.

A convolutional neural network is composed of alternatively stacked convolutional layers and spatial pooling layers. The convolutional layer is to extract feature maps by linear convolutional filters followed by nonlinear activation functions (e.g., rectifier, sigmoid, tanh, etc.). Spatial pooling is to group the local features together from spatially adjacent pixels, which is typically done to improve the robustness to slight deformations of objects. CNNs have been long studied and applied in the field of computer vision. More than a decade ago, LeCun et al. [31] trained multilayer neural networks with the back-propagation algorithm and the gradient learning technique, and then demonstrated its effectiveness on the handwritten digit recognition task. The deep CNNs have exhibited good generalization power in image-related applications. Recently, Krizhevsky et al. [32] achieved a breakthrough, outperforming the existing handcrafted features on ILSVRC 2012 which contains 1000 object classes. Since 2012, CNNs have draw a resurgence of attention in various visual recognition tasks such as image classification [32,33], semantic segmentation [34,35], object recognition [36], video analysis [37], etc. Recently the networks are going deeper, such as GoogLeNet [33] which won 2014 ILSVRC classification challenge [38] by employing 22 layers. In [39] the number of layers of the proposed residual nets reaches to 152 and achieves better performance.

There are some deep learning related works on HSI classification in the literature. Such as in [40], deep stacked autoencoders are employed to extract features. The autoencoder is a kind of unsupervised method. The proposed method in [40] combines principle component analysis (PCA), autoencoders and logistic regression, and it is not an end-to-end deep method. An end-to-end deep CNN method is proposed in [41]. The method takes the raw data as the input and outputs the predicted class labels. The number of training samples of each class is 200, and it is a relative large number.

We propose a novel CNN structure for hyperspectral image analysis, where a pixel and its neighbors in a hyperspectral image are taken as inputs of the CNN, and the final CNN output is the predicted class labels. Our designed CNN structure can be illustrated in Fig. 1. The major contributions of this work can be summarized as follows:

- Different from common visual information (e.g., RGB images), hyperspectral images can collect and process visual signals across different electromagnetic spectra. Most traditional HSI classification methods employ hand-crafted features. We design a novel CNN structure to deal with the hyperspectral image analysis problem in an end-to-end way, and the network can learn features automatically.
- Under the limitation of small training samples, we employ some network strategies (e.g., data augmentation, lager dropout rates, etc.) to alleviate the over-fitting problem in the process of learning network parameters.
- For better coping with the hyperspectral image information, we adopt the 1×1 convolutional layers to analyze the hyperspectral information, and use an average pooling layer in the whole CNNs.
- Compared with several popular features, i.e., raw spectral features, morphological features, and 3D Gabor features with traditional classifiers, the state-of-the-art classification results on three popular data sets verify the effectiveness of the proposed CNN framework. And the corresponding CNN model will be released and serve as the benchmark for the problem of hyperspectral image analysis in the research community.

The remaining part of the paper is organized as follows. The proposed network structure and design principles are presented in Section 2. The data sets used in the experiments and the evaluation methods are introduced in Section 3. Section 4 is the experimental results and analysis. The last section, Section 5, concludes the paper.

2. The CNN structure design

Since the training samples are limited, the main principle in our designed CNN structure is to alleviate the overfitting problem and gains a good generalization. We employed 1×1 convolutional kernels, improved dropout rates, discarded full connection layers, etc. The CNN structure and parameters designing are described in the following parts in detail.

2.1. Network structure

The network structure is illustrated in Fig. 1.[1] There are 3 convolutional layers in our network structure. The first two convolutional layers are followed by normalization layers and dropout layers. The input data is sent to the first convolutional layer, and the data size is $5\times 5\times N$ where N is the number of channels for hyperspectral images. In the convolutional layers, 1×1 sized kernels are employed as suggested in [42]. In the first convolutional layer, there are 128 filters. So the output of the first convolutional layer is $5\times 5\times 128$. After the convolution step, a 2×2 normalization is operated on each channel, and the next step is a dropout operation. After that the data is sent to the second convolutional layer, and also followed by a normalization layer and a dropout layer as the first convolutional layer. The output of the third convolutional layer is $5\times 5\times C$ where C is the number of classes. The global average pooling (GAP) is following the third convolutional layer. The input to GAP is the feature map with the size of $5\times 5\times C$. The GAP computes the average values on different channels, and there are C channels in this situation. So the final output of the network is a $1\times C$ vector. If the ith element in the vector has the maximal value, then i is the predicted label for the input sample. The detailed design principles of the network are described in the following subsection.

2.2. Parameter learning for the network

Data augmentation: Because the training samples are limited, the learning model tends to overfitting. To reduce overfitting in the training stage, one of the most common methods is to transform the training samples to many different ones. The transform is named as data augmentation. In our experiments, each pixel and

[1] All the source codes and model files of the proposed CNN framework will be released to the community.